

Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities

Jean-René Chazottes

Abstract We start by reviewing recent probabilistic results on ergodic sums in a large class of (non-uniformly) hyperbolic dynamical systems. Namely, we describe the central limit theorem, the almost-sure convergence to the Gaussian and other stable laws, and large deviations.

Next, we describe a new branch in the study of probabilistic properties of dynamical systems, namely concentration inequalities. They allow to describe the fluctuations of very general observables and to get bounds rather than limit laws.

We end up with two sections: one gathering various open problems, notably on random dynamical systems, coupled map lattices and so-called nonconventional ergodic averages; and another one giving pointers to the literature about moderate deviations, almost-sure invariance principle, etc.

1	Introduction	2
2	Generalities	5
3	Dynamical systems with some hyperbolicity	9
4	Limit theorems	13
5	Concentration inequalities and applications	22
6	Open questions	33
7	Notes on further results	37
	References	39

Jean-René Chazottes

Centre de Physique Théorique, CNRS-École polytechnique, 91128 Palaiseau Cedex, France, e-mail: jeanrene@cphpt.polytechnique.fr

1 Introduction

The aim of the present chapter is to roughly describe the current state of the theory of *statistical* or *probabilistic* properties of ‘chaotic’ dynamical systems. We shall restrict ourselves to *discrete-time* dynamical systems, although many of the results we review have their counterparts in flows. The basic setting is thus a state space Ω (typically a piece of \mathbb{R}^d) and a map $T : \Omega \rightarrow \Omega$. The orbit of an initial condition x_0 is the sequence of points $x_0, x_1 = Tx_0, x_2 = Tx_1, \dots$ or $\{T^k x_0; k = 0, 1, \dots\}$ (where T^k is the k -fold composition of T with itself).

The core of the probabilistic approach is the description of asymptotic *time-averages* of ‘observables’, that is, functions $f : \Omega \rightarrow \mathbb{R}$. This implies that transients become irrelevant, although transient effects may cause formidable problems in practice. The corner stone of this approach is Birkhoff’s ergodic theorem. It tells us that, *given* a measure μ left invariant by T , ‘the asymptotic *time-average* of f coincide with the space-average $\int f d\mu$ ’, except on a set of measure zero with respect to this measure. The drawback of this result is that chaotic systems typically possess uncountably many invariant ergodic measures. Is there a ‘natural’ choice ?

In this chapter, we focus on *dissipative* systems whose orbits settle on an attractor which has typically a volume (Lebesgue measure) equal to zero. In these systems, the dynamics contracts volumes but generally not in all directions: some directions may be stretched, provided some others are so much contracted that the final volume is smaller than the initial volume. This implies that, even in a dissipative system, the motion after transients may be unstable within the attractor. This instability manifests itself by an exponential separation of orbits, as time goes on, of points which initially are very close to each other on the attractor. The exponential separation takes place in the direction of stretching. Such an attractor is called *chaotic*. Of course, since the attractor is bounded, exponential separation can only hold as long as distances are small.

A famous attractor is the Hénon attractor generated by a two-dimensional map with two parameters. For some parameters, it is easy to numerically produce a ‘picture’ of the attractor. The standard way to make it is to pick ‘at random’ an initial condition in the basin of the attractor and to plot the first thousand iterates of its orbit (see Fig. 3). On the one hand, why does what is observed has something to do with the attractor since, as noticed above, it has zero volume ? On the other hand, we know that orbits of the Hénon map are not all the same: some are periodic, others are not; some come closer to the ‘turns’ than others. We also know from experience that (for a fixed T) one gets essentially the same picture independent of the choice of initial condition. Is there a mathematical explanation for this ?

These questions motivated the idea of *Sinai-Ruelle-Bowen* or SRB measures. Our computer picture can be thought as the picture of a probability measure giving mass $1/n$ to each point in an orbit of length n . Let δ_x be the point mass at x . Is there a (probability) measure μ with the property that $\frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x)} \rightarrow \mu$ for ‘most’ choices of initial conditions x , that would explain why our pictures look similar ? If such a measure does exist, it has very special properties: like all invariant probability measures, it must be supported on the attractor, but it has the peculiar ability to

influence orbits starting from various parts of the basin, including points rather far away from the support of μ . In some sense, SRB measures are the observable or physical measures.

Mathematically speaking, the theory of chaotic attractors began with the ergodic theory of *differentiable dynamical systems*, more specifically the theory of hyperbolic dynamical systems, where geometry plays a prominent role. The first systems studied in the 1960-70's were the so-called Anosov and Axiom A systems which are 'uniformly' hyperbolic and in some sense the most chaotic systems. The main results were obtained by Sinai, Ruelle and Bowen. They essentially relied on the fact that, for such systems, it is possible to construct Markov partitions enabling one to identify points in the state space with configurations in one-dimensional lattice systems of statistical mechanics [9].

The 1970's brought new outlooks and new challenges. With the aid of computer graphics, an abundance of examples showed up whose dynamics is dominated by expansions and contractions, but which do not satisfy the stringent requirements of Axiom A systems. Hénon's attractor mentioned above is a typical example. This led to a more comprehensive theory dealing with *non-uniformly* hyperbolic dynamical systems developed abstractly by Pesin and others [40, Chap. 2]. A breakthrough was made by L.-S. Young at the end of the 1990's [56, 57]. She proposed a more 'phenomenological' approach to describe in a unified framework many examples of systems with a 'localized' source of non-hyperbolicity. In particular, this provided tools to prove the existence of an SRB measure for the Hénon attractor (for a set of parameters with positive measure), see [7]. In this chapter, we shall focus on the class of systems defined by Young.

Once we know that our dynamical system (Ω, T) admits an SRB measure, we can ask for its probabilistic properties. Indeed, it can be viewed as a stationary stochastic process: the orbits (x, Tx, \dots) , where x is distributed according to μ , generate a stationary process whose finite-dimensional marginals are the measures μ_n on Ω^n given by

$$d\mu_n(x_0, \dots, x_{n-1}) = d\mu(x_0) \delta_{x_1=Tx_0} \cdots \delta_{x_{n-1}=Tx_{n-2}}.$$

This is not a product measure but the idea is that, if the system is chaotic enough, $T^k x$ is more or less independent of x provided k is large, making the process (x, Tx, \dots) behave like an independent process.

Given any observable $f : \Omega \rightarrow \mathbb{R}$, one can generate a process $\{X_n = f \circ T^n; n \geq 0\}$ on the probability space (Ω, μ) . The ergodic sum $S_n f(x) = f(x) + f(Tx) + \cdots + f(T^{n-1}x)$ is thus the partial sum of the process $\{X_n; n \geq 0\}$ and one can ask various natural questions. For instance, what is the typical size of fluctuations of $\frac{1}{n} S_n f(x)$ around $\int f d\mu$? What is the probability that $\frac{1}{n} S_n f(x)$ deviates from $\int f d\mu$ by more than some prescribed value? Does $S_n f$, appropriately renormalized, converge in law? In other words, can we prove a central limit theorem? Can we get a description of large deviations? Can we have Gaussian but also non-Gaussian limit laws? This kind of results are called *limit theorems*.

There are many quantities describing a dynamical system which can be in principle computed by observing its orbits. But the corresponding estimators are not

as simple as ergodic sums of suitably chosen observables. A prominent example (see below for details) is the periodogram which is related to the power spectrum. Therefore it is desirable to have a tool which allows to quantify fluctuations of fairly general observables for finite-length orbits. This is the scope of concentration inequalities, a new branch in the study of probabilistic theory of dynamical systems (and a quite recent branch of Probability theory as well [49]). The aim of concentration inequalities is to quantify the size of the deviations of an observable $K(x, Tx, \dots, T^{n-1}x)$ around its expectation, where $K : \Omega^n \rightarrow \mathbb{R}$ is an observable of n variables of an arbitrary expression. An ergodic sum is a very special case of such an observable and we shall see below various examples. What is imposed on K is sufficient smoothness (Lipschitz property). Depending on the ‘degree of chaos’ in the system, the deviations of K with respect to its expectation can have an extremely small probability.

From the technical viewpoint, the tool of paramount importance is the *transfer* or *Ruelle’s Perron-Frobenius operator*. This is the spectral approach to dynamical systems. We refer to book of Baladi [1] and to the lecture notes of Hennion and Hervé [41] for a throughout exposition.

Our purpose is to give a sample of recent results on the fluctuations of observables in the ergodic theory of non-uniformly hyperbolic dynamical systems. Needless to say that the overwhelming list of works in this area renders futile any attempt at an exhaustive or even comprehensive treatment within the confines of this chapter. Hopefully, this chapter provides a panoramic view of this subject. We also provide a list of directions for further research.

Before describing the contents of this chapter, a few words are in order about the bibliography. We urge the reader to consult [42] in which are gathered landmark papers illustrating the history and development of the notions of chaotic attractors and their ‘natural’ invariant measures. For numerical implementations of the theory, it is still worth reading the review paper by Eckmann and Ruelle [30]. A more recent reference, dealing both with theoretical and numerical aspects is the book by Collet and Eckmann [24]. Needless to say that the potential list of references is gigantic. Limitation of space and time forced us painfully to exclude many relevant papers. As a matter of principle, and whenever possible, we refer to the most recent articles which contain relevant pointers to the literature. We apologize for omissions.

Layout of the chapter. In Section 2 we describe the probability approach to dynamical systems and recall Birkhoff’s ergodic theorem. In Section 3 we describe the class of hyperbolic dynamical systems we will be working with. In particular, we quickly describe Young towers and SRB measures, and give several examples which will be used throughout the chapter. Section 4 is devoted to mixing (decay of correlations) and limit theorems, namely: the central limit theorem, convergence to non-Gaussian laws, exponential and sub-exponential large deviations, and convergence in law made almost sure. Section 5 is concerned with concentration inequalities and some of their applications. In Section 6 we provide a list of open problems and questions related to random dynamical systems, coupled map lattices, partially hyperbolic systems, and the Erdős-Rényi law. We end with a section where we quickly

survey results not detailed in the main text. This includes Berry-Esseen theorem, moderate deviations and the almost-sure invariance principle.

2 Generalities

We state some general definitions and recall Birkhoff's ergodic theorem.

2.1 Dynamical systems and observables

In this chapter, by 'dynamical system' we mean a deterministic dynamical system with discrete time, that is, a transformation $T : \Omega \rightarrow \Omega$ of its state space (or phase space) Ω into itself. For the sake of concreteness, one can think of Ω as a compact subset of \mathbb{R}^d . Mathematically speaking, one can deal with a compact riemannian manifold.

Every point $x \in \Omega$ represents a possible state of the system. If the system is in state x , then it will be in state $T(x)$ in the next moment of time. Given the current state $x = x_0 \in \Omega$, the sequence of states

$$x_1 = Tx_0, x_2 = Tx_1, \dots, x_n = Tx_{n-1}, \dots$$

represents the entire future or forward orbit of x_0 . We have $x_n = T^n x_0$, where T^n is the n -fold composition of T with itself. If the map T is invertible, then the past of x_0 can be determined as well ($x_{-n} = T^{-n} x_0$).

In applications, the actual states $x_n \in \Omega$ are often not observable. Instead, we usually observe the values $f(x_n)$ taken by a function f on Ω , usually called an observable. One can be thought of f as an instantaneous measurement of the system. For the sake of simplicity, we consider f to be real-valued.

More generally, we may observe the system from time 0 up to time $n - 1$ and associate to $x, Tx, \dots, T^{n-1}x$ a real number $K(x, Tx, \dots, T^{n-1}x)$. In the language of statistics, $K : \Omega^n \rightarrow \mathbb{R}$ is called an estimator. The fundamental example is the Cesàro or ergodic average of an 'instantaneous' observable $f : \Omega \rightarrow \mathbb{R}$ along an orbit up to time $n - 1$: $K_0(x, Tx, \dots, T^{n-1}x) := (f(x) + f(Tx) + \dots + f(T^{n-1}x))/n$. This is an example of an additive observable. There are many natural examples which are not as simple. An important example is the periodogram used to estimate the power spectrum of a 'signal' $\{f(x_k); k = 0, \dots, n - 1\}$. We give its definition below as well as other examples; see section 5.4.

2.2 Dynamical systems as stochastic processes

Ergodic theory is concerned with measure-preserving transformations, meaning that the map T preserves a probability measure μ on Ω : for any measurable subset $A \subset \Omega$ one has $\mu(A) = \mu(T^{-1}(A))$, where $T^{-1}(A)$ denotes the set of points mapped into A . The invariant measure μ describes the distribution of the sequence $\{x_n = T^{n-1}(x_0)\}$ for typical initial states x_0 . This vague statement is made precise by Birkhoff's ergodic theorem; see below. For a large class of non-uniformly hyperbolic systems, there is a 'natural' invariant measure, the so-called Sinai-Ruelle-Bowen measure (SRB measure for short).

A measure-preserving dynamical system is thus a probability space $(\Omega, \mathcal{B}, \mu)$ endowed with a transformation $T : \Omega \rightarrow \Omega$ leaving μ invariant. An important notion is that of an ergodic dynamical system. The invariant measure μ is said to be ergodic (with respect to T) whenever $T^{-1}(E) = E$ implies $\mu(E) = 0$ or $\mu(E) = 1$. Equivalently, ergodicity means that any invariant function $g : \Omega \rightarrow \mathbb{R}$ is μ -almost everywhere constant. That g be invariant means that $g = g \circ T$. In the measure-theoretic sense, ergodic measures are indecomposable and any invariant measure can be disintegrated into its ergodic components [44].

A measure-preserving dynamical system can be viewed as a stochastic process: the orbits (x, Tx, \dots) , where x is distributed according to μ , generate a stationary process whose finite-dimensional marginals are the measures μ_n on Ω^n given by

$$d\mu_n(x_0, \dots, x_{n-1}) = d\mu(x_0) \delta_{x_1=Tx_0} \cdots \delta_{x_{n-1}=Tx_{n-2}}.$$

This is not a product measure but the idea is that, if the system is chaotic enough, $T^k x$ is more or less independent of x provided k is large, making the process (x, Tx, \dots) behave like an independent process.

Given an observable $f : \Omega \rightarrow \mathbb{R}$, $X_k = f \circ T^k$, for each $k \geq 0$, is a random variable on the probability space $(\Omega, \mathcal{B}, \mu)$. The family $\{X_n; n \geq 0\}$ is a real-valued stationary process. The ergodic sum $S_n f(x) = f(x) + f(Tx) + \cdots + f(T^{n-1}x)$ is thus the partial sum of the process $\{X_n; n \geq 0\}$.

We shall make no attempt to define precisely what a chaotic dynamical system is. From the point of view of this chapter, we can vaguely state that it is a system such that, for sufficiently nice observables f , the process $\{f \circ T^k\}$ behave as an i.i.d.¹ process. Along the way, this crude statement will be refined.

2.3 Birkhoff's Ergodic Theorem

The fundamental theorem in ergodic theory is Birkhoff's ergodic theorem which is a far reaching generalization of Kolmogorov's strong law of large numbers for an independent process [46].

¹ i.i.d. stands for 'independent and identically distributed'

Theorem 1 (Birkhoff's ergodic theorem).

Let $(\Omega, \mathcal{B}, \mu)$ be a dynamical system and $f : \Omega \rightarrow \mathbb{R}$ be an integrable observable ($\int |f| d\mu < \infty$). Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n f(x) = f^*(x), \quad \mu - \text{almost surely and in } L^1(\mu),$$

where the function f^* is invariant ($f^* = f^* \circ T$, μ -a.s.) and such that $\int f^* d\mu = \int f d\mu$.

If the dynamical system is ergodic, then f^* is μ -almost surely a constant, whence

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n f(x) = \int f d\mu, \quad \mu - \text{almost surely.}$$

Remark 1. The previous theorem, spelled out for an integrable stationary ergodic process $\{X_n\}$, reads $n^{-1} \sum_{j=0}^{n-1} X_j \rightarrow \mathbb{E}[X_0]$ almost surely. In the non-ergodic case convergence is to the conditional expectation of X_0 with respect to the σ -algebra of invariant sets, see [46] for details.

Very often, Ω is compact and it is not difficult to show that there exists a measurable set of μ -measure one such that, in the ergodic case, $g : \Omega \rightarrow \mathbb{R}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} S_n g(x) = \int g d\mu$$

for any continuous observables. Equivalently, this means that the *empirical measure* of μ -almost every x converges towards μ in the vague (or weak-*) topology:

$$\frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x} \xrightarrow{\text{vaguely}} \mu \quad \text{almost surely.}$$

The advantage of Birkhoff's ergodic theorem is its generality. Its drawback is that a chaotic system has in general uncountably many distinct ergodic measures. Which one do we choose ? We shall see later on that the idea of a Sinai-Ruelle-Bowen measure provides an answer.

2.4 Speed of convergence and fluctuations

It is well known that not much can be said about the speed of convergence of the ergodic average to its limit in Theorem 1. First of all, one cannot know in practice if we are observing a typical orbit for which convergence indeed occurs. But even if

we knew that we have a typical orbit, it can be shown that the convergence can be arbitrarily slow (see for instance [43] for a survey).

To obtain more informations about the fluctuations of ergodic sums around their limit, we need a probabilistic formulation. Maybe the most natural question is the following:

what is the speed of convergence to zero of the probability that the ergodic average differs from its limit by more than a prescribed value ?

Formally, we want to know the speed of convergence to zero of

$$\mu \left\{ x : \left| \frac{1}{n} S_n f(x) - \int f d\mu \right| > t \right\}$$

for $t > 0$ small enough and for a large class of continuous observables f . (By Birkhoff's ergodic theorem, all what we know is that this probability goes to 0 as n goes to infinity.) In probabilistic terminology, we want to know the speed of convergence *in probability* of ergodic averages to their limit. By analogy with bounded i.i.d. processes, this speed should be exponential for 'sufficiently chaotic' systems. We shall see that it can be only polynomial when mixing is not strong enough.

Another natural issue is to determine the order of *typical values* of $S_n f - n \int f d\mu$. By analogy with a square-integrable i.i.d. process, one can expect this order to be \sqrt{n} , and, more precisely, that a central limit theorem may hold. We shall see that this is indeed the case for 'nice observables' and sufficiently chaotic systems. When chaos is 'too weak', the central limit theorem may fail and the asymptotic distribution may be non-Gaussian.

The previous issues are formulated in terms of limit theorems and concern ergodic sums. From the point of view of applications, an important problem is to estimate the probability of deviation of a general observable $K(x, Tx, \dots, T^{n-1}x)$ from its expected value. Formally, we ask if it is possible to find a positive function $b(n, t)$ such that

$$\mu \left\{ x \in \Omega : \left| K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu(y) \right| > t \right\} \leq b(n, t)$$

for any $t > 0$ and for any $n \in \mathbb{N}$, with $b(n, t)$ depending on K . When $b(n, t)$ decreases 'rapidly' with t and n , this means that $K(x, Tx, \dots, T^{n-1}x)$ is 'concentrated' around its expected value. It turns out that when the dynamical system is 'chaotic enough', this concentration phenomenon is very sharp.

To be able to answer the kind of previous questions, we shall need to make hypotheses on the dynamical systems as well as on the class of observables. Usually, Hölder continuous functions are suitable.

3 Dynamical systems with some hyperbolicity

We quickly and roughly describe the class of dynamical systems for which one can prove various probabilistic results. These systems are used to model deterministic chaos which is caused by dynamic instability, or sensitive dependence on initial conditions, together with the fact that orbits are confined in a compact region.

3.1 Hyperbolic dynamical systems

The basic model for sensitive dependence on initial conditions is that of a uniformly expanding map T on a riemannian compact manifold Ω : T is smooth and there are constants $C > 0$ and $\lambda > 1$ such that for any $x \in \Omega$ and v in the tangent space at x and for any $n \in \mathbb{N}$

$$\|DT^n(x)v\| \geq C\lambda^n\|v\|.$$

The prototypical example is $T(x) = 2x \pmod{1}$ on $\Omega = S^1$ (the unit circle), which is usually identified with the interval $[0, 1)$. The Lebesgue measure is invariant in this case.

Uniformly hyperbolic maps have the property that at each point x the tangent space is a direct sum of two subspaces E_x^u and E_x^s , one of which is expanded ($\|DT^n(x)v\| \geq C\lambda^n\|v\|$ for $v \in E_x^u$) and the other contracted ($\|DT^n(x)v\| \leq C\lambda^{-n}\|v\|$ for $v \in E_x^s$). The prototypical example is Arnold's cat map $(x, y) \mapsto (2x + y, x + y) \pmod{1}$ of the unit torus.

Non-uniform hyperbolicity refers to the fact that $C = C(x) > 0$ and $\lambda = \lambda(x) > 1$ almost-everywhere: in words, the constants depend on x and they have nice properties only on a set of full measure. For instance, the presence of a single point where $\lambda(x) = 1$ already causes important difficulties (the fundamental example being an interval map with an indifferent fixed point at 0). Another instance of loss of uniform hyperbolicity is when there is a point where the differential of T vanishes (e.g., the quadratic map or the Hénon map). A third typical situation is when the differential has discontinuities. This is the case for the Lozi map and billiards, for instance.

3.2 Attractors

We are especially interested in dissipative systems with an attractor, that is, volume-contracting maps T with an attractor Λ . By an attractor we refer to a compact invariant set with the property that all points in a neighborhood U of Λ (called its basin) are attracted to Λ (i.e. for any $x \in U$, $T^n x \rightarrow \Lambda$ as $n \rightarrow \infty$).

The prototype of a hyperbolic attractor is an Axiom A attractor. It is a smooth map T with an attractor Λ on which T is uniformly hyperbolic. These systems can

be viewed as subshifts of finite type by using a Markov partition: one can assign to each point a bi-infinite symbol sequence describing its itinerary. This sequence can be thought of as a configuration in a one-dimensional statistical mechanical system. Special measures, called SRB measures (see next section) can be constructed by pulling back adequate Gibbs measures which are invariant by the shift map; see [9] and [39, Chap. 4].

Hénon's attractor is a genuinely non-uniformly hyperbolic attractor which resisted to mathematical analysis till the 1990's.

3.3 Sinai-Ruelle-Bowen measures

We shall not define precisely Sinai-Ruelle-Bowen (SRB for short) measures but content ourselves by saying that they are the invariant measures most compatible with volume (Lebesgue measure) when volume is not preserved. Technically speaking, they have absolutely continuous conditional measures along unstable manifolds and a positive Lyapunov exponent. They provide a mechanism for explaining how local instability on attractors can produce coherent statistics for orbits starting from large sets in the basin. In particular, an SRB measure μ is 'observable' in the following sense: there exists a subset V of the basin of attraction with positive Lebesgue measure such that for any continuous observable f on Ω and any initial state $x \in V$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int f d\mu,$$

or, more compactly

$$\frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x} \xrightarrow{\text{vaguely}} \mu.$$

The point of this property is that the set of 'good states' has positive Lebesgue measure although the measure μ is concentrated on the attractor which has zero Lebesgue measure. (Notice that this property does not follow from Birkhoff's ergodic theorem.)

For one-dimensional maps, absolutely continuous invariant measures (with respect to Lebesgue measure) are examples of SRB measures.

Roughly speaking, the approach to non-uniformly hyperbolic systems of L.-S. Young, which will be sketched below, can be considered as 'phenomenological' in the sense that it aims at modeling concrete dynamical behaviors observed in various examples. An 'axiomatic approach' can be followed which seeks to relax the conditions that define Axiom A systems in the hope of systematically enlarging the set of maps with SRB measures. For an account on this second approach, we refer to [40, Chap. 2]. For a nice and non-technical survey on SRB measures, we recommend reading [58].

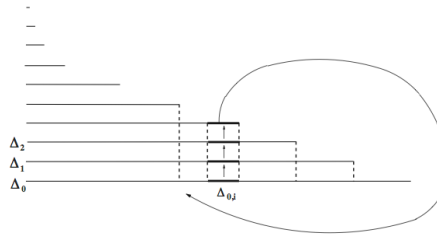
3.4 Dynamical systems modeled by a Young tower

In the 1970s, many examples were numerically observed whose dynamics are dominated by expansions and contractions but which do not meet the stringent requirements of Axiom A systems. The most famous example is likely the Hénon mapping which displays a ‘strange attractor’ for certain parameters. Such examples remained mathematically intractable until the 1990’s.

L.-S. Young developed a general scheme to study the probabilistic properties of a class of ‘predominantly hyperbolic’ dynamical systems, including the Hénon attractor and other famous examples. Very roughly the picture is as follows. The general set up is that $T : \Omega \curvearrowright$ is a nonuniformly hyperbolic system in the sense of Young [56, 57] with a return time function R that decays either exponentially [56], or polynomially [57]. In particular, $T : \Omega \curvearrowright$ is modeled by a Young tower constructed over a ‘uniformly hyperbolic’ base $Y \subset \Omega$. The degree of non-uniformity is measured by the return time function $R : Y \rightarrow \mathbb{Z}^+$ to the base.

More precisely, by a classical construction in ergodic theory, one can construct from (Y, T^R) an extension (Δ, F) , called a Young tower in the present setting. In particular, there exists a continuous map $\pi : \Delta \rightarrow \Omega$ such that $\pi \circ F = T \circ \pi$. In general π need not be one-to-one or onto. One can visualize a tower by writing that $\Delta = \bigcup_{\ell=0}^{\infty} \Delta_{\ell}$ where Δ_{ℓ} can be identified with the set $\{x \in Y : R(x) > \ell\}$, that is, the ℓ -th floor of the tower. In particular, Δ_0 is identified with Y . The dynamics in the tower is as follows: each point $x \in \Delta_0$ moves up the tower until it reaches the top level above x , after which it returns to Δ_0 , see Fig. 1. Moreover, F has a countable Markov partition $\{\Delta_{\ell,j}\}$ with the property that π maps each $\Delta_{\ell,j}$ injectively onto Y , which has a hyperbolic product structure. Each of the local unstable manifolds defining the product structure of $\pi(\Delta_0)$ meet $\pi(\Delta_0)$ in a set of positive Lebesgue measure. Further analytic and regularity conditions are imposed. We shall not give further details and refer the reader to [56, 57] and [19].

Fig. 1 Schematic representation of the tower map $F : \Delta \curvearrowright$.



Systems modeled by Young towers are more flexible than Axiom A systems in that they are permitted to be non-uniformly hyperbolic: roughly speaking, think of uniform hyperbolicity as required only for the return map to the base. Reasonable singularities and discontinuities are also allowed: they do not appear in Y . As we shall see, a number of probabilistic properties of $T : \Omega \curvearrowright$ are actually captured by

the tail properties of R . The basic result proved in [56, 57] is the following, where m^u denotes Lebesgue measure on unstable manifolds.

Theorem 2. *Let $T : \Omega \curvearrowright$ be a dynamical system modeled by a Young tower. If $\int R dm^u < \infty$, then T has an ergodic SRB measure. If $\gcd\{R_i\} = 1$, there is a unique SRB measure denoted by μ .*

Of course, $\int R dm^u = \sum_{n \geq 1} m^u\{R > n\}$. In the sequel, we shall implicitly assume that $\gcd\{R_i\} = 1$, without loss of generality.

3.5 Some examples

The best known example of a non-uniformly expanding map of the interval is the so-called Maneville-Pomeau map modelling intermittency. It is expanding except at 0 where the slope of the map is one (neutral fixed point). For the sake of definiteness², consider the map

$$T_\alpha(x) = \begin{cases} x + 2^\alpha x^{1+\alpha} & \text{if } x \in [0, 1/2) \\ 2x - 1 & \text{if } x \in [1/2, 1] \end{cases} \quad (1)$$

where $\alpha \in (0, 1)$ is a parameter. It is well-known that there is a unique absolutely continuous invariant measure $d\mu(x) = h(x)dx$ and $h(x) \sim x^{-\alpha}$ as $x \rightarrow 0$. There is a Young tower with base $Y = [1/2, 1]$ and $\text{Leb}\{y \in Y : R(y) > n\} = \mathcal{O}(n^{-1/\alpha})$.

Another fundamental one-dimensional example is given by the quadratic family $T_a : [-1, 1] \curvearrowright$ with $T_a(x) = 1 - ax^2$, where $a \in [1, 2]$, and for which 0 is a critical point (the slope vanishes). For a set of parameters of positive Lebesgue measure, this map preserves a unique absolutely continuous probability measure. Its density has an inverse square-root singularity. In this example, one can construct a tower map with a return-time function which has an exponentially decreasing tail.

An important example of a dynamical system in the plane modeled by a Young tower with a return time decaying exponentially is the Lozi map:

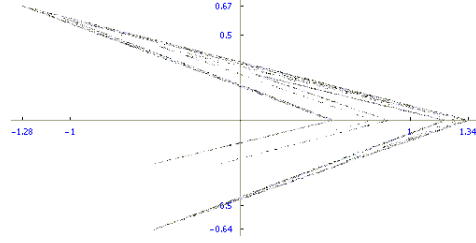
$$T_{a,b} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 - a|x| + y \\ bx \end{pmatrix}$$

which possesses an attractor depicted in Fig. 2. Lozi's map is much simpler to analyse than the famous Hénon map:

$$T_{a,b} : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix}$$

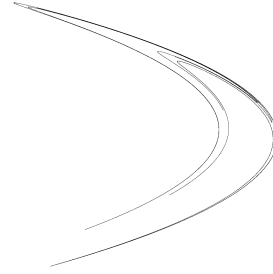
² The explicit formula (1) is not important, what matters is only the local behavior around the fixed point.

Fig. 2 Simulation of the Lozi attractor for $a = 1,7$ and $b = 0,5$.



For certain parameters, this map has an attractor displayed in Fig. 3. For the so-called

Fig. 3 Simulation of the Hénon attractor for $a = 1,4$ and $b = 0,3$. Notice that the existing results do not cover these ‘historical’ values.



Benedicks-Carleson parameters³, it is possible to prove [7] that the Hénon attractor fits the general scheme of Young towers with exponential tails. In particular, there is a unique SRB measure whose support is the attractor.

Important examples of maps, which are conservative, are billiard maps, like planar Lorentz gases and Sinai’s billiard. They can be also modeled by Young towers. We refer to [56] but also to [21] for a conceptual account avoiding technicalities.

4 Limit theorems

In this section we review some limit theorems obtained for the class of systems previously described.

4.1 Covariance and decay of correlations

Definition 1 (Correlations).

For a dynamical system (Ω, T, μ) and an observable $f : \Omega \rightarrow \mathbb{R}$ in $L^2(\mu)$, the *autocovariance* of order $\ell \geq 0$ of the process $\{f \circ T^k; k \geq 0\}$ is defined as

³ These parameters form a subset of \mathbb{R}^2 with positive Lebesgue measure [5].

$$C_f(\ell) := \int f \cdot f \circ T^\ell d\mu - \left(\int f d\mu \right)^2.$$

More generally, for a pair f, g of observables in $L^2(\mu)$, the covariance of order ℓ of the processes $\{f \circ T^k; k \geq 0\}$ and $\{g \circ T^k; k \geq 0\}$ is defined as

$$C_{f,g}(\ell) := \int f \cdot g \circ T^\ell d\mu - \int f d\mu \int g d\mu.$$

In dynamical systems, it is customary to call the auto-covariance of order ℓ the “*correlation coefficient*” of order ℓ .

The auto-covariance, or more generally, the covariance, is the basic indicator of a chaotic behavior: for large values of ℓ , the random variables f and $f \circ T^\ell$ should be nearly independent, *i.e.* the coefficient $C_f(\ell)$ should decay to 0 as ℓ grows. Two factors affect the rapidity of this decay: the strength of chaos in the underlying dynamical system $T : \Omega \rightarrow \Omega$ and the regularity of the observable f .

Recall that a dynamical system $(\Omega, \mathcal{B}, T, \mu)$ is mixing if for any two measurable sets $A, B \subset \Omega$ one has $\mu(A \cap T^{-n}B) \xrightarrow{n \rightarrow \infty} \mu(A)\mu(B)$. It is easy to prove that the system is mixing if and only if correlations decay, *i.e.*, $C_{f,g}(\ell) \xrightarrow{n \rightarrow \infty} 0$ for every pair of $f, g \in L^2(\mu)$.

The speed or rate of the decay of correlations (also called the rate of mixing) is crucial in the statistical analysis of chaotic systems.

Theorem 3 (Mixing and decay of correlations [56, 57, 55, 35]).

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower and μ its SRB measure. The system is mixing and the rate of decay of correlations for Hölder continuous observables is directly related to the behavior of $m^u\{R > n\}$ as $n \rightarrow \infty$.

- For example, if $m^u\{R > n\} = \mathcal{O}(e^{-an})$ for some $a > 0$, then (T, μ) has exponential decay of correlations.
- If $m^u\{R > n\} = \mathcal{O}(1/n^\gamma)$ for some $\gamma > 1$, then (T, μ) has polynomial decay of correlations. More precisely, $C_f(\ell) = \mathcal{O}(1/\ell^{\gamma-1})$.

For the Hénon map with Benedicks-Carleson parameters, correlations for Hölder continuous observables decay exponentially fast. The intermittent map (1) has polynomial decay of correlations: $C_f(\ell) = \mathcal{O}(1/\ell^{\frac{1}{\alpha}-1})$. Two-dimensional examples with an intermittent behavior come from billiards. Chernov and Zhang studied in [22, 23] several classes of billiards for which the decay of correlations is $\mathcal{O}((\log \ell)^c / \ell^{1/\alpha-1})$ for some parameter α taking values in $(0, 1/2]$.

4.2 Central limit theorem

We start by a definition.

Definition 2 (Central limit theorem).

Let (Ω, T, μ) be a dynamical system and $f : \Omega \rightarrow \mathbb{R}$ an observable in $L^2(\mu)$. We say that f satisfies the central limit theorem (CLT for short) with respect to (T, μ) if there exists $\sigma_f \geq 0$ such that

$$\lim_{n \rightarrow \infty} \mu \left\{ x : \frac{S_n f(x) - n \int f d\mu}{\sqrt{n}} \leq t \right\} = \frac{1}{\sqrt{2\pi}\sigma_f} \int_{-\infty}^t e^{-\frac{u^2}{2\sigma_f^2}} du, \quad \forall t \in \mathbb{R}. \quad (2)$$

In probabilistic notation, the previous convergence is written compactly as

$$\frac{S_n f - n \int f d\mu}{\sqrt{n}} \xrightarrow{\text{law}} \mathcal{N}_{0, \sigma_f^2},$$

where $\mathcal{N}_{0, \sigma_f^2}$ stands for the Gaussian law with mean 0 and variance σ_f^2 .

When $\sigma_f = 0$ the right-hand side has to be understood as the Heaviside function.

In probabilistic terms, this definition asks for the *convergence in law* of the ergodic average ‘zoomed out’ by the factor \sqrt{n} to a random variable whose law is $\mathcal{N}_{0, \sigma_f^2}$.

By analogy with i.i.d. processes, one expects that σ_f be the variance of the process $\{f \circ T^n\}$. If it were an i.i.d. process, we would have

$$\sigma_f^2 = \text{Var}(S_n f / \sqrt{n}) = \int f^2 d\mu - \left(\int f d\mu \right)^2 d\mu = C_f(0),$$

where $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$ is the variance of X . But because of the correlations between f and $f \circ T^n$, this is not the case. A natural candidate for the variance is

$$\sigma_f^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \int (S_n f - n \int f d\mu)^2 d\mu,$$

provided the limit exists. Simple algebra, using the invariance of μ under T , gives

$$\frac{1}{n} \int (S_n f - n \int f d\mu)^2 d\mu = C_f(0) + 2 \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} C_f(\ell).$$

It is simple to prove that if

$$\sum_{j=1}^{\infty} |C_f(j)| < \infty$$

then

$$\lim_{n \rightarrow \infty} \sum_{\ell=1}^{n-1} \frac{n-\ell}{n} C_f(\ell) = \sum_{\ell=1}^{\infty} C_f(\ell),$$

whence

$$\sigma_f^2 = C_f(0) + 2 \sum_{\ell=1}^{\infty} C_f(\ell). \quad (3)$$

We have the following theorem.

Theorem 4 (Central limit theorem, [56, 57]).

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower and μ its SRB measure. Let $f : \Omega \rightarrow \mathbb{R}$ be a Hölder continuous observable. If $\int R^2 d\mu < \infty$ (which implies $\sum_{\ell \geq 1} |C_f(\ell)| < \infty$), then f satisfies the central limit theorem with respect to (T, μ) .

For the class of systems discussed in this paper, it is well-known that typically $\sigma_f^2 > 0$. Indeed, $\sigma_f^2 = 0$ only for Hölder observables lying in a closed subspace of infinite codimension.

For example, Hölder continuous observables satisfy the CLT for the Hénon map with Benedicks-Carleson parameters. For the map (1), the CLT holds if $\alpha < 1/2$. We shall see what happens when $\alpha \geq 1/2$ later on.

There are examples of convergence to the Gaussian law but with a non-classical renormalizing sequence $(\sqrt{n \log n})$, instead of (\sqrt{n}) . This is the case for Bunimovich's billiard (stadium) where correlations decay only as $1/n$ (where n is the number of collisions); see [2].

In essence, the central limit theorem tells us that *typically* (i.e. with very high probability),

$$S_n f - n \int f d\mu = \mathcal{O}(\sqrt{n}).$$

In other words, the typical fluctuations of $S_n f/n$ around $\int f d\mu$ are of order $1/\sqrt{n}$. But, in principle, $S_n f$ can take values as large as n , i.e. $S_n f/n - \int f d\mu$ can be of order one, but with a small probability. Such fluctuations are naturally called 'large deviations'. This is the subject of the next section.

4.3 Large deviations

For a bounded i.i.d. process $\{X_n\}$, it is a classical result in probability, usually called Cramér's theorem [27], that $\mathbb{P}\{|n^{-1}(X_0 + \dots + X_{n-1}) - \mathbb{E}[X_0]| > \delta\}$ decays exponentially with n . Moreover,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left\{ \left| \frac{X_0 + \dots + X_{n-1}}{n} - \mathbb{E}[X_0] \right| > \delta \right\} = -\mathbf{I}(\delta).$$

Typically, the function \mathbf{I} (the so-called *rate function*) is strictly convex and vanishes only at 0^4 (hence it is non-negative). Since the process is bounded, its domain is a finite interval. The rate function turns out to be the Legendre transform of the cumulant generating function $\theta \mapsto \log \mathbb{E}[\exp(\theta X_0)]$.

One expects this exponential decay for the probability of deviation in ‘sufficiently chaotic’ dynamical systems and for a Hölder continuous observable f . For notational convenience, assume that $\int f d\mu = 0$. The goal is to prove that there exists a rate function $\mathbf{I}_f : \mathbb{R} \rightarrow [0, +\infty]$ such that

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mu \left\{ x \in \Omega : \frac{1}{n} S_n f(x) \in [a - \varepsilon, a + \varepsilon] \right\} = -\mathbf{I}_f(a).$$

In many situations, such a result is obtained by proving that the cumulant generating function

$$\Psi_f(z) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \int e^{z S_n f} d\mu$$

exists and is smooth enough for z real in an interval containing the origin. Then the rate function is the Legendre transform of Ψ_f . However, as we shall see, when chaos is not strong enough, one may indeed get subexponential decay rates for large deviations (and therefore there is no rate function).

For systems modeled by a Young tower with exponential tails, we have the following result. It turns out that the logarithmic moment generating function $\Psi_f(z)$ can be studied for complex z .

Theorem 5 (Cumulant generating functions [54, 51]).

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower and μ its SRB measure. Assume that $m^n\{R > n\} = \mathcal{O}(e^{-an})$ for some $a > 0$. Let $f : \Omega \rightarrow \mathbb{R}$ be a Hölder continuous observable such that $\int f d\mu = 0$.

- Then there exist positive numbers $\eta = \eta(f)$ and $\xi = \xi(f)$ such that the logarithmic moment generating function Ψ_f exists and is analytic in the strip

$$\{z \in \mathbb{C} : |\operatorname{Re}(z)| < \eta, |\operatorname{Im}(z)| < \xi\}.$$

- In particular, $\Psi_f'(0) = \int f d\mu$ and $\Psi_f''(0) = \sigma_f^2$, which is the variance (3) of the process $\{f \circ T^n\}$. Moreover, $\Psi_f(z)$ is strictly convex for real z provided $\sigma_f^2 > 0$.

From this kind of result, one can deduce the following result by using Gartner-Ellis theorem or the like (see [27, section 4.5] and [41, pp. 102–103]). Notice that it is enough for Ψ_f to be differentiable to apply this theorem.

⁴ The rate function must vanish at 0 in view of Birkhoff’s ergodic theorem.

Theorem 6 (Exponential large deviations [51, 54]).

Under the same assumptions as in the previous theorem, let \mathbf{I}_f be the Legendre transform of Ψ_f , i.e. $\mathbf{I}_f(t) = \sup_{z \in (-\eta, \eta)} \{tz - \Psi_f(z)\}$. Then for any interval $[a, b] \subset [\Psi'_f(-\eta), \Psi'_f(\eta)]$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu \left\{ x \in \Omega : \frac{1}{n} S_n f(x) \in [a, b] \right\} = - \inf_{t \in [a, b]} \mathbf{I}_f(t).$$

Remark 2. Using a general theorem of Bryc [11], one can deduce the central limit theorem from Theorem 5. We stress that analyticity of Ψ_f is necessary. In general, if Ψ_f is only C^∞ (ensuring that $\Psi''_f(0) = \sigma_f^2$), it is false that the central limit theorem follows from exponential large deviations.

We now turn to systems modeled by a Young tower with sub-exponential tails. In this case, there is no rate function and one gets sub-exponential large deviation bounds.

Theorem 7 (Sub-exponential large deviations [50]).

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower and μ its SRB measure. Assume that $m^u\{R > n\} = \mathcal{O}(1/n^\gamma)$ for some $\gamma > 1$. Let $f : \Omega \rightarrow \mathbb{R}$ be a Hölder continuous observable such that $\int f d\mu = 0$. Then, for any $\varepsilon > 0$

$$\mu \left\{ x \in \Omega : \left| \frac{1}{n} S_n f(x) \right| > \varepsilon \right\} \leq \frac{C_{f, \varepsilon}}{n^{\gamma-1}}, \quad \text{for any } n \in \mathbb{N}.$$

Notice that according to Theorem 3, the decay is the same as that for correlations. The dependence in ε of the constant $C_{f, \varepsilon}$ is in ε^{-2q} where $q > \max(1, \gamma - 1)$.

Let us again use our favorite example, namely the Manneville-Pomeau map, to illustrate the preceding result. In this case, one can also prove a lower bound for the probability of large deviations. Indeed, for the map (1), the theorem applies with $\gamma = \frac{1}{\alpha}$, where $\alpha \in (0, 1)$. Recall that for $\alpha \in (0, 1/2)$, the central limit theorem holds (see Section 4.2), but it fails when $\alpha \in [1/2, 1)$ (See Section 4.4 below).

Moreover, it is proved in [50] that there is a nonempty open set of Hölder observables f for which $n^{-\frac{1}{\alpha}+1}$ is a lower bound for large deviations for n sufficiently large. For these observables, we have for any $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \frac{\log \mu \left\{ x \in [0, 1] : \left| \frac{1}{n} S_n f(x) \right| > \varepsilon \right\}}{\log n} = -\frac{1}{\alpha} + 1. \quad (4)$$

4.4 Convergence to non-Gaussian laws

The purpose of this section is to show what happens when the CLT fails but one still has convergence in law, but with a renormalizing sequence different from (\sqrt{n}) . For the reader's convenience, we recall the notion of domain of attraction for an observable and a classical theorem about stable laws for i.i.d. processes.

A function f , defined on a probability space (Ω, \mathcal{B}, m) , is said to *belong to a domain of attraction* if it fulfills one the following three conditions:

- I. It belongs to $L^2(\Omega)$.
- II. One has $\int \mathbb{1}_{\{|f|>x\}} dm \sim x^{-2} \ell(x)$, for some function ℓ such that $L(x) := 2 \int_1^x \frac{\ell(u)}{u} du$ is of slow variation and unbounded.
- III. There exists $p \in (1, 2)$ such that

$$\int \mathbb{1}_{\{f>x\}} dm = (c_1 + o(1))x^{-p}L(x) \quad \text{and} \quad \int \mathbb{1}_{\{f<-x\}} dm = (c_2 + o(1))x^{-p}L(x),$$

where c_1, c_2 are nonnegative real numbers such that $c_1 + c_2 > 0$, and L is of slow variation.

Note that the three conditions are mutually exclusive.

The above definition of domain of attraction is motivated by the following well-known, classical result in Probability (see *e.g.* [33]):

Theorem 8 (Convergence to stable laws for i.i.d. processes).

Let Z be a random variable belonging to a domain of attraction. Let Z_0, Z_1, \dots be a sequence of independent, identically distributed, random variables with the same law as Z . In all cases, we set $A_n = n\mathbb{E}[Z]$ and

- 1. if condition I holds, we set $B_n = \sqrt{n}$ and $\mathcal{W} = \mathcal{N}_{0, \mathbb{E}[Z^2] - \mathbb{E}[Z]^2}$;
- 2. if condition II holds, we let B_n be a renormalizing sequence with $nL(B_n) \sim B_n^2$, and $\mathcal{W} = \mathcal{N}_{0,1}$;
- 3. if condition III holds, we let B_n be a renormalizing sequence such that $nL(B_n) \sim B_n^p$. Define $c = (c_1 + c_2)\Gamma(1-p)\cos(\frac{p\pi}{2})$ and $\beta = \frac{c_1 - c_2}{c_1 + c_2}$.

Let $\mathcal{W} = \mathcal{W}_{p,c,\beta}$ be the law with characteristic function

$$\mathbb{E}[e^{it\mathcal{W}}] = e^{-c|t|^p(1-i\beta \operatorname{sgn}(t)\tan(\frac{p\pi}{2}))}, \quad (1 < p \leq 2, c > 0, |\beta| \leq 1). \quad (5)$$

Then

$$\frac{\sum_{i=0}^{n-1} Z_i - A_n}{B_n} \xrightarrow{\text{law}} \mathcal{W}.$$

The case $p = 2$ corresponds to the Gaussian law. For $p < 2$, the corresponding distributions are said to have ‘heavy tails’ since $\mathbb{P}\{Z > x\} = (c_1 + o(1))x^{-p}$ and $\mathbb{P}\{Z < -x\} = (c_2 + o(1))x^{-p}$. The conditions put on the distribution of Z are almost necessary and sufficient to get a convergence in law of that type, we only restricted the range of p ’s, which could also be taken in the interval $(0, 1]$.

We illustrate the occurrence of non-Gaussian limit laws in the most important example, that is, the Pomeau-Manneville map (1).

Theorem 9 (Convergence to stable laws for the Manneville-Pomeau map [35]).

Let T_α be the map of the interval (1), with $\alpha \in (0, 1)$ and μ its unique absolutely continuous, invariant, probability measure. Let $f : [0, 1] \rightarrow \mathbb{R}$ be a Hölder observable and assume that $\int f d\mu = 0$.

- If $\alpha < 1/2$ then the central limit theorem holds (this is a special instance of Theorem 4).
- If $\alpha > 1/2$ then:
 - if f is Lipschitzian and $f(0) = 0$, then the central limit theorem holds;
 - if $f(0) \neq 0$ then $\frac{1}{n^\alpha} S_n f$ converges in law to the stable law $\mathcal{W}_{\frac{1}{\alpha}, c, \text{sgn}(f(0))}$ whose characteristic function is given by (5).

When $\alpha = 1/2$ and $f(0) \neq 0$, there is convergence to the Gaussian law but with the unusual renormalizing sequence $(\sqrt{n \log n})$ (instead of \sqrt{n}). See [34] for more details.

4.5 Convergence in law made almost sure

The aim of this section is to show that whenever we can prove a limit theorem in the classical sense for a dynamical system, we can prove a suitable almost-sure version based on an empirical measure with log-average.

The prototype of such a theorem is the almost-sure central limit theorem: if X_n is an i.i.d. L^2 sequence with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = 1$, then, almost surely,

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{\sum_{j=0}^{k-1} X_j / \sqrt{k}} \xrightarrow{\text{law}} \mathcal{N}_{0,1} \quad (6)$$

where “ $\xrightarrow{\text{law}}$ ” means weak convergence of probability measures on \mathbb{R} . Here and henceforth, δ_x is the Dirac mass at x . This result should be compared to the classical central limit theorem, which can be stated as follows:

$$\mathbb{E}[\mathbb{1}_{\{\sum_{j=0}^{n-1} X_j/\sqrt{n} \leq t\}}] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du$$

for any $t \in \mathbb{R}$. To better compare these theorems, it is worth noticing that (6) implies that *almost surely*

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \mathbb{1}_{\{\sum_{j=0}^{k-1} X_j/\sqrt{k} \leq t\}} \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-u^2/2} du \quad (7)$$

for any $t \in \mathbb{R}$. So, instead of taking the expected value, we take a logarithmic average and obtain an almost-sure convergence.

In fact, whenever there is independence and a classical limit theorem, the corresponding almost-sure limit theorem also holds (under minor technical conditions), see [8] and references therein.

Let us put the following general definition:

Definition 3 (Almost sure limit theorem towards a random variable).

Let S_n be a sequence of random variables on a probability space, and let B_n be a renormalizing sequence.⁵ We say that S_n/B_n satisfies an almost sure limit theorem towards a law \mathcal{W} if, for almost all ω ,

$$\frac{1}{\log N} \sum_{k=1}^N \frac{1}{k} \delta_{S_k(\omega)/B_k} \xrightarrow{\text{law}} \mathcal{W}.$$

We now turn to the dynamical system context. The almost-sure central limit theorem, for instance, takes the form

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{S_k f(x)/\sqrt{k}} \xrightarrow{\text{law}} \mathcal{N}_{0, \sigma_f^2}, \text{ for } \mu - \text{almost every } x,$$

where, for notational simplicity, we assume that $\int f d\mu = 0$.

In the paper [18], we proved that “whenever we can prove a limit theorem in the classical sense for a dynamical system, we can prove a suitable almost-sure version”. More precisely, we investigated three methods that are used to prove limit theorems in dynamical systems: spectral methods, martingale methods, and induction arguments. We showed that whenever these methods apply, the corresponding limit theorem admits a suitable almost-sure version.

For instance, one has the following result.

Theorem 10 (Convergence in law made almost sure [18]).

Let $T : \Omega \curvearrowright$ be a dynamical system modeled by a Young tower and let μ its

⁵ A renormalization function is a function $B : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ of the form $B(x) = x^d L(x)$ where $d > 0$ and L is a normalized slowly varying function. The corresponding renormalizing sequence is $B_n := B(n)$.

SRB measure. Let $f : \Omega \rightarrow \mathbb{R}$ be a Hölder continuous observable such that $\int f d\mu = 0$. Then, if

$$\mu \left\{ x \in \Omega : \frac{S_k f(x)}{B_k} \leq t \right\} \xrightarrow[k \rightarrow \infty]{} \mathcal{W}((-\infty, t])$$

for every $t \in \mathbb{R}$ at which \mathcal{W} is continuous, for a certain law \mathcal{W} and for a certain renormalizing sequence (B_n) , then

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{S_k f / B_k} \xrightarrow{\text{law}} \mathcal{W} \quad \mu - \text{almost-surely.}$$

Let us illustrate this theorem with a few examples. For any dynamical system modeled by a Young tower with L^2 tails, one has

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{S_k f(x) / \sigma_f \sqrt{k}} \xrightarrow{\text{law}} \mathcal{N}_{0, \sigma_f^2}.$$

For the Manneville-Pomeau map (1), this is true for $\alpha \in (0, 1/2)$. When $\alpha > 1/2$, this is still the case provided that $f(0) = 0$ and f is Lipschitz. If $f(0) \neq 0$, then

$$\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{S_k f(x) / k^\alpha} \xrightarrow{\text{law}} \mathcal{W}_{\frac{1}{\alpha}, c, \text{sgn}(f(0))}$$

(see Theorem 9).

5 Concentration inequalities and applications

5.1 Introduction

We start by the simplest occurrence of the concentration of measure phenomenon [49]. Consider an independent sequence of Bernoulli random variables $(\eta_i)_{0 \leq i \leq n-1}$ (i.e. $\mathbb{P}(\eta_i = -1) = \mathbb{P}(\eta_i = 1) = 1/2$, whence $\mathbb{E}[\eta_i] = 0$). Then one has the following classical inequality (Chernov's bound):

$$\mathbb{P} \left(\left| \sum_{i=0}^{n-1} \eta_i \right| \geq t \right) \leq 2 \exp \left(- \frac{t^2}{2n} \right), \quad \forall t \geq 0. \quad (8)$$

This exponential inequality reflects the most important theorem of probability, imprecisely stated as follows: “In a long sequence of tossing a fair coin, it is likely that heads will come up nearly half of the time.” Indeed, if we let B_n be the number of

1's in the sequence $(\eta_i)_{0 \leq i \leq n-1}$, then $\sum_{i=0}^{n-1} \eta_i = 2B_n - n$, and so (8) is equivalent to

$$\mathbb{P}\left(\left|B_n - \frac{n}{2}\right| \geq t\right) \leq 2 \exp\left(\frac{-2t^2}{n}\right), \quad \forall t \geq 0.$$

This is of course a much stronger statement than the Strong Law of Large Numbers.

The perspective of concentration inequalities is to look at the random variable $Z_n = \sum_{i=0}^{n-1} \eta_i$ as a function of the individual variables η_i . Inequality (8), when Z_n is normalized by n (since it can take values as large as n) can be phrased pretty offensively by saying that

$$\frac{Z_n}{n} \text{ is essentially constant } (= 0).$$

The scope of concentration inequalities is to understand to what extent a general function K of n random variables X_0, \dots, X_{n-1} , and not just the sum of them, concentrates around its expectation like a sum of Bernoulli random variables. Of course, the smoothness of K has to play a role, as well as the dependence between the X_i 's.

Stated informally as a principle, the measure of concentration phenomenon is the following:

“A random variable that smoothly depends on the influence of many weakly dependent random variables is, on the appropriate scale, very close to a constant.”

This statement is of course quantified by statements like (8) or weaker ones, as we shall see.

In the context of dynamical systems, there are many examples of random variables $K(X_0, \dots, X_{n-1})$ which appear naturally but are defined in an indirect or complicated way. Concentration inequalities, when available, allow to obtain, in a systematic way, *a priori* bounds on the fluctuations of $K(X_0, \dots, X_{n-1})$ around its expectation by using a simple information on K , namely its Lipschitz constants.

5.2 Concentration inequalities: abstract definitions

We formulate some abstract definitions.

Let Ω be a metric space. A real-valued function K on Ω^n is separately Lipschitz if, for any i , there exists a constant $\text{Lip}_i(K)$ such that

$$\left|K(x_0, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{n-1}) - K(x_0, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_{n-1})\right| \leq \text{Lip}_i(K) d(x_i, x'_i)$$

for all points $x_0, \dots, x_{n-1}, x'_i$ in Ω .

Consider a stationary process $\{X_0, X_1, \dots\}$ taking values in Ω .

Definition 4 (Exponential concentration inequality).

We say that the process $\{X_0, X_1, \dots\}$ satisfies an exponential concentration inequality if there exists a constant $C > 0$ such that, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$, one has

$$\mathbb{E} \left[e^{K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})]} \right] \leq e^{C \sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2}. \quad (9)$$

In some cases, it is not reasonable to hope for such a strong inequality. This leads to the following definition.

Definition 5 (Polynomial concentration inequality).

We say that the process $\{X_0, X_1, \dots\}$ satisfies a polynomial concentration inequality with moment $p \geq 2$ if there exists a constant $C > 0$ such that, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$, one has

$$\mathbb{E} [|K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})]|^p] \leq C \left(\sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2 \right)^{p/2}. \quad (10)$$

An important special case of (10) is for $p = 2$, which gives an inequality for the variance of $K(X_0, \dots, X_{n-1})$:

$$\text{Var}(K(X_0, \dots, X_{n-1})) \leq C \sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2. \quad (11)$$

After these definitions, a few comments are in order.

- The crucial point in (9) and (10) is that the constant C does not depend neither on K nor on n . It solely depends on the process.
- These inequalities are not asymptotic, they hold true for any n .
- Obviously (9) is a much stronger inequality than (10). For instance, one can get (11) from (9) as follows: Multiply K by $\lambda \neq 0$, subtract 1 from both sides, divide by λ^2 ; conclude by using Taylor expansion and by letting λ go to 0.
- An important consequence of the previous inequalities is a control on the deviation probabilities of $K(X_0, \dots, X_{n-1})$ from its expectation:
If a stationary process $\{X_n\}$ satisfies the exponential concentration inequality (9) then, for any $t > 0$, one has

$$\mathbb{P} \{ |K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})]| > t \} \leq 2 e^{-\frac{t^2}{4C \sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2}}. \quad (12)$$

If the process satisfies the polynomial concentration inequality (10), one gets that for any $t > 0$

$$\mathbb{P} \{ |K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})]| > t \} \leq C t^{-q} \left(\sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2 \right)^{q/2}. \quad (13)$$

To prove (12), we use Markov's inequality and (9): for any $t, \lambda > 0$

$$\begin{aligned}
& \mathbb{P} \{K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})] > t\} \\
&= \mathbb{P} \left\{ \exp \left(\lambda (K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})]) \right) > \exp(\lambda t) \right\} \\
&\leq e^{-\lambda t} \mathbb{E} \left[e^{\lambda (K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})])} \right] \\
&\leq e^{-\lambda t} e^{C\lambda^2 \sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2}.
\end{aligned}$$

This upper bound is minimized when $\lambda = t / (2C \sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2)$, whence

$$\mathbb{P} \{K(X_0, \dots, X_{n-1}) - \mathbb{E}[K(X_0, \dots, X_{n-1})] > t\} \leq e^{-\frac{t^2}{4C \sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2}}.$$

The previous procedure is usually called the ‘Chernoff bounding trick’. Of course, we can apply this inequality to $-K$ and deduce at once (12).

Inequality (13) follows immediately from Markov's inequality. \square

5.3 Concentration inequalities for dynamical systems

We now present concentration inequalities in the setting of non-uniformly hyperbolic dynamical systems. In a forthcoming paper with S. Gouëzel [19] we prove the following theorems. Let us notice that we take separately Lipschitz observables for the sake of simplicity. All results are valid in the Hölder case (see [19, Section 7.1]).

5.3.1 Main results

Theorem 11 (Exponential concentration inequality [19]).

Let (Ω, T, μ) be a dynamical system modeled by a Young tower with exponential tails. Then it satisfies an exponential concentration inequality: there exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$,

$$\int e^{K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu(y)} d\mu(x) \leq e^{C \sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2} \quad (14)$$

As a consequence of the Chernoff bounding trick (see the previous section), we get, for any $t > 0$ and for any $n \in \mathbb{N}$,

$$\mu \left\{ x \in \Omega : K(x, Tx, \dots, T^{n-1}x) - \int K(y, \dots, T^{n-1}y) d\mu(y) > t \right\} \leq e^{-\frac{t^2}{4C \sum_{j=0}^{n-1} \text{Lip}_j(K)^2}}. \quad (15)$$

The same bound holds for lower deviations by applying (15) to $-K$.

There are well-known dynamical systems (X, T) which can be modeled by a Young tower with exponential tails [56]. Examples of invertible dynamical systems fitting this framework are for instance Axiom A attractors, Hénon's attractor for Benedicks-Carleson parameters [7], piecewise hyperbolic maps like the Lozi attractor, some billiards with convex scatterers, etc. A non-invertible example is the quadratic family for Benedicks-Carleson parameters.

Theorem 12 (Polynomial concentration inequality [19]).

Let (Ω, T, μ) be a dynamical system modeled by a Young tower. Assume that, for some $q \geq 2$, $\int R^q d\mu^u < \infty$. Then it satisfies a polynomial concentration inequality with moment $2q - 2$, i.e., there exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$,

$$\int \left| K(x, Tx, \dots, T^{n-1}x) - \int K(y, Ty, \dots, T^{n-1}y) d\mu(y) \right|^{2q-2} d\mu(x) \leq C \left(\sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2 \right)^{q-1}. \quad (16)$$

As a direct application of Markov's inequality, we get from that, for any $t > 0$ and for any $n \in \mathbb{N}$,

$$\mu \left\{ x \in \Omega : \left| K(x, Tx, \dots, T^{n-1}x) - \int K(y, \dots, T^{n-1}y) d\mu(y) \right| > t \right\} \leq C \frac{(\sum_{\ell=0}^{n-1} \text{Lip}_\ell(K)^2)^{q-1}}{t^{2q-2}} \quad (17)$$

For the Manneville-Pomeau map, we know that the exponential concentration inequality cannot be true. Indeed, (4) is clearly an obstruction. Applying Theorem 12, we get a concentration inequality with moment Q for any $Q < \frac{2}{\alpha} - 2$ when $\alpha \in (0, 1/2)$. Applying (13) yields a deviation bound in $n^{-\frac{1}{\alpha}+1+\delta}$, for any $\delta > 0$. This is very close to the upper bound in $n^{-\frac{1}{\alpha}+1}$ guaranteed by Theorem 7. In fact, one can get an optimal deviation inequality and get the latter bound, but we need the notion of a weak polynomial concentration inequality that we do not want to detail here, see [19].

5.3.2 About the literature

The first paper in which a concentration inequalities was proved for dynamical systems is [25]: an exponential concentration inequality is established for piecewise uniformly expanding maps of the interval. For dynamical systems (X, T) modeled by a Young tower with exponential tails, a polynomial concentration inequality with moment 2 (variance) was proved in [16]. Regarding systems with subexponential decay of correlations, the first result was obtained in [15] for the Manneville-Pomeau map (1): a polynomial concentration inequality with moment 2 was proved for $\alpha \leq 4 - \sqrt{15}$. The above theorems, proved in [19], improve all these results in several ways.

5.4 A sample of applications of concentration inequalities

We present some applications of concentration inequalities to show them in action. Some more, as well as all proofs, can be found in [17, 18, 20, 25].

5.4.1 Warming-up with ergodic sums

Let us apply the exponential inequality to the basic example is $K_0(x_0, \dots, x_{n-1}) = f(x_0) + \dots + f(x_{n-1})$ where f is a Lipschitz observable. We obviously have $\text{Lip}_i(K_0) = \text{Lip}(f)$ for any $i = 0, \dots, n-1$. When evaluated along an orbit segment $x, \dots, T^{n-1}x$, we of course get the ergodic sum $S_n f(x)$. Assuming that (15) holds one gets

$$\mu \left\{ x \in \Omega : \left| \frac{1}{n} S_n f(x) - \int f d\mu \right| > t \right\} \leq 2e^{-\frac{nt^2}{4\text{CLip}(f)^2}}, \forall t > 0.$$

Compared with large deviations (see Subsection 4.3), we observe that this is the right order in n . The large deviation result provides a much more accurate description of this deviation probability as $n \rightarrow \infty$. But the previous inequality shows how small this deviation probability is already for finite n 's.

5.4.2 Correlations

Let (Ω, T, μ) be an ergodic dynamical system and $f : \Omega \rightarrow \mathbb{R}$ be a Lipschitz observable such that $\int f d\mu = 0$. An obvious estimator of the correlation coefficient $C_f(k)$ (cf. Def. 1) is

$$\widehat{C}_f(n, k, x) = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) f(T^{j+k} x).$$

Indeed, an immediate consequence of Birkhoff's ergodic theorem is that

$$\widehat{C}_f(n, k, x) \xrightarrow[n \rightarrow \infty]{} C_f(k), \quad \mu - \text{a.s.}$$

Observe that $\int \widehat{C}_f(n, k, x) d\mu = C_f(k)$ by the invariance of the measure.

We have the following result.

Theorem 13 (Correlation coefficients).

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower and μ its SRB measure. Let $f : \Omega \rightarrow \mathbb{R}$ be a Lipschitz observable such that $\int f d\mu = 0$.

- If the tower has exponential tails, there exists $D > 0$ such that for any $t > 0$ and any $k, n \in \mathbb{N}$

$$\mu \left\{ x \in \Omega : \left| \widehat{C}_f(n, k, x) - C_f(k) \right| > t \right\} \leq 2e^{-D \frac{n^2 t^2}{n+k}}.$$

- If, for some $q \geq 2$, $\int R^q dm^u < \infty$, then there exists $G > 0$ such that for any $t > 0$ and any $k, n \in \mathbb{N}$

$$\mu \left\{ x \in \Omega : \left| \widehat{C}_f(n, k, x) - C_f(k) \right| > t \right\} \leq G \left(\frac{n+k}{n^2} \right)^{q-1} \frac{1}{t^{2q-2}}.$$

The proof is easy. One considers the function

$$K(x_0, \dots, x_{n+k-1}) = \frac{1}{n} \sum_{j=0}^{n-1} f(x_j) f(x_{j+k})$$

of $n+k$ variables. It is obvious that $\text{Lip}_i(K) \leq \|f\|_\infty \text{Lip}(f)/n$. Applying (12) and (13) yields the desired inequality.

5.4.3 Empirical measure

Let (Ω, T, μ) be an ergodic dynamical system. Birkhoff's ergodic theorem (see Subsection 2.3) implies that the empirical measure $\mathcal{E}_n(x) = (1/n) \sum_{j=0}^{n-1} \delta_{T^j x}$ converges vaguely to μ . We want to obtain a 'speed' for this convergence, so we need to define a distance. We use the Kantorovich distance dist_K . For two probability measures μ_1 and μ_2 on Ω , it is defined as

$$\text{dist}_K(\mu_1, \mu_2) = \sup \left\{ \int g d\mu_1 - \int g d\mu_2 : g : \Omega \rightarrow \mathbb{R} \text{ is } 1\text{-Lipschitz} \right\}.$$

This distance is compatible with the vague topology.

We are led to consider the observable

$$K(x, Tx, \dots, T^{n-1}x) = \text{dist}_K(\mathcal{E}_n(x), \mu).$$

Theorem 14 (Empirical measure).

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower with exponential tails and μ its SRB measure. Then, for any $t > 0$ and for any $n \in \mathbb{N}$

$$\mu \left\{ x \in \Omega : \left| \text{dist}_K(\mathcal{E}_n(x), \mu) - \int \text{dist}_K(\mathcal{E}_n(y), \mu) d\mu(y) \right| > \frac{t}{\sqrt{n}} \right\} \leq 2e^{-t^2/4C}.$$

This theorem follows at once from (12) and the fact that the function K defined above has all its Lipschitz constants bounded by $1/n$. A natural step further is to try to get an upper bound for $\int \text{dist}_K(\mathcal{E}_n(\cdot), \mu) d\mu$. There is no general good bound in general; one has first to restrict to one-dimensional systems (because there is a special representation for the Kantorovich distance in terms of the distribution functions). Second, the regularity of the observables for which there is exponential decay of correlations is crucial. We mention only one result for the quadratic map $T_a(x) = 1 - ax^2$ acting on $\Omega = [-1, 1]$, where $a \in [0, 2]$. For Benedicks-Carleson parameters, we mentioned above that this system can be modeled by a Young tower with exponential tails. In fact there is an exponential decay of correlations for more general observables than the Hölder ones, namely for observables with bounded variation [59]. This allows to prove that

$$\int \text{dist}_K(\mathcal{E}_n(\cdot), \mu) d\mu \leq \frac{B}{\sqrt{n}}$$

for some $B > 0$. Hence we deduce the following result from (14).

Theorem 15.

Consider the map $T_a(x) = 1 - ax^2$ acting on $\Omega = [-1, 1]$ for a in the Benedicks-Carleson set of parameters. Then there exist $D, t_0 > 0$ such that for any $t \geq t_0$ and for any $n \in \mathbb{N}$

$$\mu \left\{ x \in \Omega : \text{dist}_K(\mathcal{E}_n(x), \mu) > \frac{t}{\sqrt{n}} \right\} \leq 2e^{-Dt^2}.$$

A natural question is to estimate the density of the absolutely continuous invariant measure of a one-dimensional dynamical system. A classical estimator is the so-called kernel density estimator. We refer to [17, 19] for details and results.

5.4.4 Tracing orbits

We use concentration inequalities to quantify the tracing properties of some subsets of orbits. The basic problem can be formulated as follows. Let A be a set of initial conditions and x an initial condition not in A : How well can one approximate the orbit of x by an orbit from an initial condition of A ? One can measure the ‘average quality of tracing’ by defining

$$\mathcal{S}_A(x, n) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} d(T^j x, T^j y)$$

where d is the distance on Ω . Assume that $\text{diam}(\Omega) = 1$. We have the following result.

Theorem 16.

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower with exponential tails and μ its SRB measure. There exists a constant $c > 0$ such that for any subset $A \subset X$ with strictly positive μ -measure, for any $n \in \mathbb{N}$ and for any $t > 0$

$$\mu \left\{ x \in \Omega : \mathcal{S}_A(x, n) > c \frac{\sqrt{\log n}}{\mu(A)\sqrt{n}} + \frac{t}{\sqrt{n}} \right\} \leq e^{-t^2/4C}$$

(where $C > 0$ is the constant appearing in Theorem 11).

Proof. The function of n variables

$$K(x_0, \dots, x_{n-1}) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} d(x_j, T^j y).$$

is separately Lipschitz and it is easy to check that $\text{Lip}_i(K) \leq 1/n$ for any $i = 0, \dots, n-1$. We use (12) to get at once

$$\mu \left\{ x : \mathcal{S}_A(x, n) > \int \mathcal{S}_A(y, n) d\mu(y) + \frac{t}{\sqrt{n}} \right\} \leq e^{-t^2/4C}. \quad (18)$$

We now estimate $\int \mathcal{S}_A(y, n) d\mu(y)$ from above. Fix $s > 0$ and define the set

$$B_s = \left\{ x : \mathcal{S}_A(x, n) > \int \mathcal{S}_A(y, n) d\mu(y) + \frac{s}{\sqrt{n}} \right\}.$$

We have the identity

$$\int \mathcal{S}_A(y, n) d\mu(y) = \int_A \mathcal{S}_A(y, n) d\mu(y) + \int_{A^c \cap B_s^c} \mathcal{S}_A(y, n) d\mu(y) + \int_{B_s} \mathcal{S}_A(y, n) d\mu(y).$$

The first integral is equal to 0 by the very definition of \mathcal{S}_A . The second one is bounded by

$$\left(\int \mathcal{S}_A(y, n) d\mu(y) + \frac{s}{\sqrt{n}} \right) \mu(A^c).$$

And the third one is bounded by $\mu(B_s)$ because $\mathcal{S}_A(y, n) \leq 1$. By (18) one has

$$\mu(B_s) \leq e^{-s^2/4C}.$$

Hence

$$\int \mathcal{S}_A(y, n) d\mu(y) \leq \left(\int \mathcal{S}_A(y, n) d\mu(y) + \frac{s}{\sqrt{n}} \right) \mu(A^c) + e^{-s^2/4C},$$

i.e.

$$\int \mathcal{S}_A(y, n) d\mu(y) \leq \mu(A)^{-1} \left(\frac{s}{\sqrt{n}} + e^{-s^2/4C} \right).$$

To finish the proof, it remains to optimize over $s > 0$. □

For a system modeled by a Young tower with polynomial tails, one can obtain a weaker bound, see [19].

5.4.5 Integrated periodogram

Let (Ω, T, μ) be an ergodic dynamical system and $f : \Omega \rightarrow \mathbb{R}$ be a Lipschitz observable with $\int f d\mu = 0$. Define the empirical integrated periodogram of the process $\{f \circ T^k\}$ by

$$\mathcal{J}_n(x, \omega) = \int_0^\omega \frac{1}{n} \left| \sum_{j=0}^{n-1} e^{-ijs} f(T^j x) \right|^2 ds, \quad \omega \in [0, 2\pi].$$

Let

$$\mathcal{J}(\omega) = C_f(0)\omega + 2 \sum_{k=1}^{\infty} \frac{\sin(\omega k)}{k} C_f(k),$$

that is, the cosine Fourier transform of the sequence of correlation coefficients. (Recall that $C_f(k) = \int f \cdot f \circ T^k d\mu$.) One can prove the following theorem.

Theorem 17.

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower with exponential tails and μ its SRB measure. Let $f : \Omega \rightarrow \mathbb{R}$ be a Lipschitz function such that $\int f d\mu = 0$. There exist some positive constants c_1, c_2 such that for any $n \in \mathbb{N}$ and for any $t > 0$

$$\mu \left\{ x \in \Omega : \sup_{\omega \in [0, 2\pi]} |\mathcal{J}_n(x, \omega) - \mathcal{J}(\omega)| > t + \frac{c_1(1 + \log n)^{3/2}}{\sqrt{n}} \right\} \leq e^{-c_2 n t^2 / (1 + \log n)^2}.$$

The proof can be found in [19].

5.4.6 Almost-sure central limit theorem

We come back to the almost-sure central limit theorem (cf. Subsection 4.5). Let f be a Lipschitz observable such that $\int f d\mu = 0$. For convenience, let

$$\mathcal{A}_n = \frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \delta_{S_k f / \sqrt{k}}.$$

This is a random measure on \mathbb{R} . Given $x \in \Omega$, $\mathcal{A}_n(x)$ is a measure. To measure its closeness to the Gaussian law $\mathcal{N}_{0, \sigma_f^2}$, we use the Kantorovich distance dist_K . For two probability measures μ_1 and μ_2 on \mathbb{R} , it is defined as

$$\text{dist}_K(\mu_1, \mu_2) = \sup \left\{ \int g d\mu_1 - \int g d\mu_2 : g : \mathbb{R} \rightarrow \mathbb{R} \text{ is } 1\text{-Lipschitz} \right\}.$$

Convergence in this distance entails both weak convergence and convergence of the first moment.

Theorem 18 (Almost-sure central limit theorem).

Let $T : \Omega \rightarrow \Omega$ be a dynamical system modeled by a Young tower such that

$$\int R^2 dm^\mu < \infty$$

and μ its SRB measure. Let $f : \Omega \rightarrow \mathbb{R}$ be a Lipschitz observable with $\int f d\mu = 0$. Assume that $\sigma_f^2 > 0$. Then

$$\text{dist}_K(\mathcal{A}_n(x), \mathcal{N}_{0, \sigma_f^2}) \rightarrow 0 \quad \text{for } \mu\text{-a.e. } x \in \Omega.$$

This is slightly stronger than the usual almost-sure central limit theorem. In fact, a more general statement is true: if a process $\{X_k\}$ satisfies the central limit theorem and (11), then the previous theorem is true. This is the way it is proved in [17].

6 Open questions

In this section, we list various open questions. The list we present is by no means exhaustive.

6.1 Random dynamical systems

In order to model the effect of noise on a discrete-time dynamical system, it is natural to introduce models obtained by compositions of different maps rather than by repeated applications of exactly the same transformation. The idea is to study sequences of maps ‘picked at random’ in some stationary fashion. We refer to [40, Chap. 5] for a survey.

The simplest case is the following. We assume that the phase space is contained in \mathbb{R}^d and that there is a sequence of i.i.d., \mathbb{R}^d -valued, random variables ξ_0, ξ_1, \dots such that, instead of observing the orbit of the initial condition x , one observes sequences $\{x_n\}$ of points in the state space given by

$$x_{n+1} = T(x_n) + \varepsilon \xi_n$$

where ε is a fixed parameter (the amplitude of the noise if $|\xi_n|$ is of order one). The process $\{x_n\}$ is called a *stochastic perturbation* of the dynamical system T . By construction, it is a one-parameter family of Markov chains. If we assume that ξ_n has a density ρ with respect to Lebesgue measure, the transition probability of the chain is given by

$$p(x_{n+1}|x_n) = \frac{1}{\varepsilon} \rho\left(\frac{x_{n+1} - T(x_n)}{\varepsilon}\right).$$

One expects that in the limit $\varepsilon \rightarrow 0$ (the zero-noise limit), the right-hand side converges to $\delta(x_{n+1} - T(x_n))$ and that, if μ_ε is an invariant measure for the chain, then its accumulation points (in vague topology) should be invariant measures for T . There are reasons to believe that under fairly general conditions, SRB measures may be natural candidates for zero-noise limits, hence they should be stochastically stable. This is indeed proved for Axiom A systems and certain non-uniformly hyperbolic systems, see *e.g.* [26] and [6] for the Hénon map.

A natural question is to prove concentration inequalities for random dynamical systems, in particular for the additive noise model. This would lead, for instance, to quantitative informations on the distance between the empirical measure of the process $\{x_n\}$ and the SRB measure μ as a function of n and ε .

The above setting concerns ‘dynamical noise’. Another relevant situation is ‘observational noise’: one observes the process $y_n = x_n + \varepsilon \xi_n$ and the goal is merely to extract $\{x_n\}$, and eventually try to reconstruct T [48].

6.2 Coupled map lattices

Coupled map lattices are a class of (discrete-time) spatially extended dynamical systems which were introduced in the 1980's by physicists. We refer to the lecture notes [13] for more details and background.

The basic set-up is a state space $\Omega = I^{\mathbb{Z}^d}$ where $I \subset \mathbb{R}$ is a compact interval, typically $[0, 1]$. There is a 'local' dynamics $\tau : I \rightarrow I$ which defines an 'unperturbed' dynamics T_0 on Ω by $(T_0(x))_i = \tau(x_i)$, $i \in \mathbb{Z}^d$. Then one defines a perturbed dynamics by introducing couplings $\Phi_\varepsilon : \Omega \rightarrow \Omega$ of the form $\Phi_\varepsilon(x) = x + A_\varepsilon(x)$. The basic (and most studied) example is the 'diffusive' nearest neighbor coupling

$$(\Phi_\varepsilon(x))_i = x_i + \frac{\varepsilon}{2d} \sum_{|i-j|=1} (x_j - x_i), \quad i \in \mathbb{Z}^d.$$

Of course, ε measures the strength of the coupling. The dynamics we are interested in is

$$T_\varepsilon := \Phi_\varepsilon \circ T_0.$$

The study of such dynamical systems offer many challenges and a lot of questions remain open [13].

From the point of view of probabilistic properties, the following is known, see [3] and references therein. The local map τ on the unit interval I is assumed to be continuous and piecewise C^2 . The expansion rate is assumed to be bigger than 2: $|\tau'| > 2$ and both the first- and second-order derivatives are bounded. The couplings are assumed to be diffusive and of finite range (the above example corresponds to a range equal to one). Under these conditions, the coupled map lattice T_ε has a unique observable measure μ_ε in the sense that, for $m^{\otimes \mathbb{Z}^d}$ -almost every point $x \in \Omega$ state,

$$\frac{1}{n} \sum_{k=0}^{n-1} \delta_{T_\varepsilon^k x} \xrightarrow{\text{vaguely}} \mu_\varepsilon.$$

This measure is exponentially mixing both in time and space. Moreover, any Lipschitz function on $I^{\mathbb{Z}^d}$ depending on a finite number of coordinates satisfies the central limit theorem with respect to $(T_\varepsilon, \mu_\varepsilon)$. The authors also prove a local limit theorem. All these results hold provided that ε is small enough. As the authors point out, their tools also allow to prove exponential large deviations.

A natural question is to prove concentration inequalities in this context. One expects an exponential concentration inequality to hold.

6.3 Partially hyperbolic systems

As mentioned above, the theory of hyperbolic dynamical systems initially developed from the notion of uniform hyperbolicity. This notion can be weakened in essentially two ways. One of these is to retain hyperbolicity without uniformity, which leads to the theory of non-uniformly hyperbolic dynamical systems. The class of systems modeled by Young towers described in this chapter is an important subclass of such systems.

The other generalization is to retain uniformity without hyperbolicity by allowing a center direction in which any expansion or contraction is in a uniform way slower than the expansion and contraction in the unstable and stable subspaces. Such systems are called *partially hyperbolic*. Among the basic examples are time-one maps of Anosov flows (the center direction is the flow direction), quasi-hyperbolic toral automorphisms and mostly contracting diffeomorphisms. We refer to [40, Chap. 1] for a survey.

In [29], the author proves many probabilistic results such as the central limit theorem (and its refinements like the almost-sure invariance principle) and exponential large deviations.

It would be nice to establish concentration inequalities for partially hyperbolic systems.

6.4 Nonconventional ergodic averages

Nonconventional or multiple ergodic averages are typically of the form

$$\frac{1}{n} \sum_{k=0}^{n-1} f_1(T^k x) f_2(T^{2k} x) \cdots f_\ell(T^{\ell k} x).$$

That is, one considers the averages of products of, say, bounded measurable functions along an arithmetic progression of length ℓ for an arbitrary integer $\ell \geq 1$. The case $\ell = 1$ is of course the standard case. Such averages originated in the ergodic theoretic proof by Furstenberg of Szemerédi's theorem on arithmetic progressions based on the so-called multiple recurrence theorem [32]. For a dynamical system (X, T, μ) which is weakly mixing, the above averages converge in L^2 to $\prod_{k=1}^{\ell} \int f_k d\mu$.

The next questions are about fluctuations of nonconventional averages when the f_j 's are, say, Lipschitz functions : central limit theorem, large deviations and concentration properties. Regarding the central limit theorem, a first step was done by Kifer [45] for uniformly hyperbolic systems (for averages along more general progressions). Large deviations seem much more difficult to analyse and turn out to be nontrivial even for i.i.d. processes (see [12]).

A transfer operator approach remains to be introduced to tackle such problems because the usual machinery does not seem appropriate. Remarkably, concentration

inequalities, if available for the system at hand, apply straightforwardly and provide nontrivial informations while they ‘ignore’ the fine structure of these averages. We leave as an exercise to the reader the derivation of such concentration bounds.

6.5 Erdős-Rényi law for nonuniformly hyperbolic systems and applications to multifractal analysis

We come back to large deviations (see Subsection 4.3). When a rate function does exist for a dynamical system (see Theorem 6), the following question is natural:

given an observable f , is it possible to extract the rate function \mathbf{I}_f solely from a typical orbit of the system ?

With a different motivation, this question was answered by Erdős and Rényi [31] in the context of i.i.d. random variables. In the context of dynamical systems, the first result was obtained in [14] for a class of piecewise, uniformly expanding maps of the interval. For this class, Theorem 6 is valid and one can in fact get refined large deviation estimates necessary to obtain the following result. Given an observable f and t in the domain of \mathbf{I}_f , let

$$M_k(x) = \max \{ S_k f(T^j x) : 0 \leq j \leq \lfloor \exp(k\mathbf{I}_f(t)) \rfloor - k \}$$

In words, we are looking for the largest ergodic sum of f in a window of width k inside the orbit of x up to time $\lfloor \exp(k\mathbf{I}_f(t)) \rfloor - k$.

Theorem 19 (Erdős-Rényi law for uniformly expanding maps of the interval [14]).

Let $T : [0, 1] \rightarrow [0, 1]$ be a piecewise C^2 , uniformly expanding map which is topologically mixing and μ its unique absolutely continuous invariant measure. Let $f : [0, 1] \rightarrow \mathbb{R}$ be an observable of bounded variation⁶. Then, there exists $t^* > 0$ such that, for any $|t| \leq t^*$ and for Lebesgue-almost every $x \in [0, 1]$

$$\lim_{k \rightarrow \infty} \frac{M_k(x)}{k} = t.$$

More precisely, one has almost everywhere

$$\limsup_{k \rightarrow \infty} \frac{M_k(x) - kt}{\log k} \leq \frac{1}{2u}$$

and

$$\liminf_{k \rightarrow \infty} \frac{M_k(x) - kt}{\log k} \geq -\frac{1}{2u},$$

where $u = \mathbf{I}'_f(t)$.

Notice that this theorem gives an optimal rate of convergence, the same as in the i.i.d. case obtained by Deheuvels *et al.* (see [14]).

In view of Theorem 6 and the technique used in [14], one expects that Theorem 19 be true for systems modeled by a Young tower with exponential tails. This was partially showed in [28], but only in the one-dimensional case, and with a non-optimal rate.

On the side of applications, Theorem 6 allows to construct an estimator for \mathbf{I}_f . This is particularly relevant to the estimation of multifractal spectra, see [4].

7 Notes on further results

We quickly describe or barely mention other results that we could not develop in the main text.

7.1 More on the central limit theorem

It is natural to ask for a speed of convergence in the central limit theorem. This type of result is called a *Berry-Esseen theorem*.

For systems modeled by a Young tower with exponential tails, one has the following. Let $f : \Omega \rightarrow \mathbb{R}$ be a Hölder continuous observable. Assume that $\sigma_f > 0$. Then there exists a constant $c = c(f) > 0$ such that

$$\sup_{t \in \mathbb{R}} \left| \mu \left\{ x : \frac{S_n f(x) - n \int f d\mu}{\sqrt{n}} \leq t \right\} - \frac{1}{\sqrt{2\pi}\sigma_f} \int_{-\infty}^t e^{-\frac{u^2}{2\sigma_f^2}} du \right| \leq \frac{c}{\sqrt{n}}, \quad \forall n \in \mathbb{N}.$$

The speed of convergence can be slower. Let us again illustrate this by looking at the map T_α given by (1). For $0 < \alpha < 1/2$ and f Hölder continuous (which is not of the form $g - g \circ T_\alpha$), we know that the central limit theorem holds (see end of Section 4.2).

- If $0 < \alpha < 1/3$ then one gets a speed of order $\mathcal{O}(1/\sqrt{n})$ as above.
- If $1/3 < \alpha < 1/2$ and $f(0) \neq 0$, the speed is $\mathcal{O}(1/n^{\frac{1}{2\alpha}-1})$.

We refer the interested reader to [36] for more details and proofs, where a ‘local limit theorem’ is also proved.

7.2 Moderate deviations

One can also characterize the fluctuations of $S_n f$ which are of an order intermediate between \sqrt{n} (central limit theorem) and n (large deviations). Such fluctuations, when suitably scaled, satisfy large deviations type estimates with a quadratic rate function determined by σ_f^2 . We have the following theorem:

Theorem 20 (Moderate deviations [54]).

Let $T : \Omega \curvearrowright$ be a dynamical system modeled by a Young tower and μ its SRB measure. Assume that $m^u\{R > n\} = \mathcal{O}(e^{-an})$ for some $a > 0$. Let $f : \Omega \rightarrow \mathbb{R}$ be a Hölder continuous observable which is not of the form $g - g \circ T$ (whence $\sigma_f^2 > 0$). Let a_n be an increasing sequence of positive real numbers such that $\lim_{n \rightarrow \infty} a_n/\sqrt{n} = \infty$ and $\lim_{n \rightarrow \infty} a_n/n = 0$. Then for any interval $[a, b] \subset \mathbb{R}$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{a_n^2/n} \log \mu \left\{ x \in \Omega : \frac{S_n f(x) - n \int f d\mu}{a_n} \in [a, b] \right\} = - \inf_{t \in [a, b]} \frac{t^2}{2\sigma_f^2}.$$

For the case of systems modeled by Young towers with polynomial tails, see [50].

7.3 Far beyond the CLT: the invariance principle

The almost sure invariance principle is a very strong reinforcement of the central limit theorem: it ensures that the trajectories of a process can be matched with the trajectories of a Brownian motion in such a way that almost surely the error between the trajectories is negligible compared to the size of the trajectory.

For $\lambda \in (0, 1/2]$ and Σ^2 a (possibly degenerate) symmetric semi-positive-definite $d \times d$ matrix, we say that an \mathbb{R}^d -valued process (A_0, A_1, \dots) satisfies an almost sure invariance principle with error exponent λ and limiting covariance Σ^2 if there exist a probability space \mathcal{P} and two processes (A_0^*, A_1^*, \dots) and (B_0, B_1, \dots) on \mathcal{P} such that:

1. the processes (A_0, A_1, \dots) and (A_0^*, A_1^*, \dots) have the same distribution;
2. the random variables (B_0, B_1, \dots) are independent and distributed as $\mathcal{N}_{0, \Sigma^2}$;
3. and almost surely in \mathcal{P}

$$\left| \sum_{\ell=0}^{n-1} A_\ell^* - \sum_{\ell=0}^{n-1} B_\ell \right| = o(n^\lambda).$$

A Brownian motion at integer times coincides with a sum of i.i.d. Gaussian variables, hence this definition can also be formulated as an almost sure approximation by a Brownian motion, with error $o(n^\lambda)$.

In the dynamical system context, take $A_\ell = f \circ T^\ell$ where $f : \Omega \rightarrow \mathbb{R}^d$ is regular. It is proved in [52] by martingale methods and then in [37] with purely spectral methods, that a dynamical systems modeled by Young towers satisfy the almost-sure invariance principle. Namely, this is the case if $\int R^q dm^u < \infty$ for $q > 2$ and for observables $f : \Omega \rightarrow \mathbb{R}^d$ which are Hölder continuous. The relevance of considering \mathbb{R}^d -valued observable is that, for instance, the position variable of the planar periodic Lorentz gas with finite horizon approximates a two-dimensional Brownian motion.

The almost-sure invariance principle implies in particular the central limit theorem, the functional central limit theorem, and the law of iterated logarithm, among others, see e.g. [38, 53]. It also implies the almost-sure central limit theorem [47].

Acknowledgements The author thanks Sébastien Gouëzel for useful comments. He also thanks Cesar Maldonado and Mike Todd for a careful reading.

References

1. Baladi, V.: Positive transfer operators and decay of correlations. Advanced Series in Nonlinear Dynamics **16**. World Scientific, 2000
2. Bálint, P., Gouëzel, S.: Limit theorems in the stadium billiard. Comm. Math. Phys. **263** (2006), no. 2, 461512

3. Bardet, J.-B., Gouëzel, S., Keller, G.: Limit theorems for coupled interval maps. *Stoch. Dyn.* **7** (2007), no. 1, 1736
4. Beck, C., Touchette, H.: Nonconcave entropies in multifractals and the thermodynamic formalism. *J. Stat. Phys.* **125** (2006), no. 2, 459475
5. Benedicks, M., Carleson, L.: The dynamics of the Hénon map. *Ann. of Math. (2)* **133** (1991), no. 1, 73169
6. Benedicks, M., Viana, M.: Random perturbations and statistical properties of Hénon-like maps. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **23** (2006), no. 5, 713752
7. Benedicks, M., Young, L.-S.: Markov extensions and decay of correlations for certain Hénon maps. In: *Géométrie complexe et systèmes dynamiques* (Orsay, 1995). *Astérisque* **261** (2000), 13–56
8. Berkes, I., Csáki, E.: A universal result in almost sure central limit theory. *Stochastic Process. Appl.* **94** (2001), no. 1, 105–134
9. Bowen, R.: Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Second revised edition. With a preface by David Ruelle. Edited by J.-R. Chazottes. *Lecture Notes in Mathematics*, 470. Springer-Verlag, Berlin, 2008
10. Brockwell, P., Davis, R.: Time series: theory and methods. Second edition. *Springer Series in Statistics*. Springer-Verlag, New York, 1991
11. Bryc, W.: A remark on the connection between the large deviation principle and the central limit theorem. *Statist. Probab. Lett.* **18** (1993), no. 4, 253–256
12. Carinci, G., Chazottes, J.-R., Giardinà, C., Redig, F.: Nonconventional averages along arithmetic progressions and lattice spin systems. Preprint (2012), arXiv:1110.2354.
13. Dynamics of coupled map lattices and of related spatially extended systems. Lectures from the school-forum (CML 2004) held in Paris, June 21–July 2, 2004. Edited by J.-R. Chazottes and B. Fernandez. *Lecture Notes in Physics*, 671. Springer, Berlin, 2005
14. Chazottes, J.-R., Collet, P.: Almost sure central limit theorems and Erdős-Rényi type law for expanding maps of the interval. *Ergod. Th. Dynam. Sys.* **25** (2005), no. 2, 419–441
15. Chazottes, J.-R., Collet, P., Redig, F., Verbitskiy, E.: A concentration inequality for interval maps with an indifferent fixed point. *Ergod. Th. Dynam. Sys.* **29** (2009), 1097–1117
16. Chazottes, J.-R., Collet, P., Schmitt, B.: Devroye inequality for a class of non-uniformly hyperbolic dynamical systems. *Nonlinearity* **18** 2323–2340 (2005)
17. Chazottes, J.-R., Collet, P., Schmitt, B.: Statistical consequences of Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. *Nonlinearity* **18** 2341–2364 (2005)
18. Chazottes, J.-R., Gouëzel, S.: On almost-sure versions of classical limit theorems for dynamical systems. *Probab. Th. & Rel. Fields* **138** (2007), no. 1-2, 195–234
19. Chazottes, J.-R., Gouëzel, S.: Optimal concentration inequalities for dynamical systems. To appear in *Commun. Math. Phys.* (2012)
20. Chazottes, J.-R., Maldonado, C.: Concentration bounds for entropy estimation in one-dimensional Gibbs measures. *Nonlinearity* **24** (2011), no. 8, 2371–2381
21. Chernov, N., Young, L. S.: Decay of correlations for Lorentz gases and hard balls. In *Hard ball systems and the Lorentz gas*, 89–120, *Encyclopaedia Math. Sci.*, 101, Springer (2000)
22. Chernov, N., Zhang, H.-K.: Billiards with polynomial mixing rates. *Nonlinearity* **18** (2005), no. 4, 15271553
23. Chernov, N., Zhang, H.-K.: Improved estimates for correlations in billiards. *Comm. Math. Phys.* **277** (2008), no. 2, 305321
24. Collet, P., Eckmann, J.-P.: Concepts and results in chaotic dynamics: a short course. *Theoretical and Mathematical Physics*. Springer (2006)
25. Collet, P., Martínez, S., Schmitt, B.: Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. Theor. Rel. Fields* **123**, 301–322 (2002)
26. Cowieson, W., Young, L.-S.: SRB measures as zero-noise limits. *Ergodic Theory Dynam. Systems* **25** (2005), no. 4, 11151138
27. Dembo, A., Zeitouni, O.: Large deviations techniques and applications. Corrected reprint of the second (1998) edition. *Stochastic Modelling and Applied Probability*, 38. Springer (2010)

28. Denker, M., Nicol, M.: Erdős-Rényi laws for hyperbolic dynamical systems, preprint of Mittag-Leffler Institute (2010)
29. Dolgopyat, D.: Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.* **356** (2004), no. 4, 1637–1689
30. Eckmann, J.-P., Ruelle, D.: Ergodic theory of chaos and strange attractors. *Rev. Modern Phys.* **57** (1985), no. 3, part 1, 617656
31. Erdős, P., Rényi, A.: On a new law of large numbers. *J. Analyse Math.* **23** 1970 103–111.
32. Furstenberg, H.: *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton Univ. Press (1981)
33. Gnedenko, B. V., Kolmogorov, A. N.: *Limit distributions for sums of independent random variables*. Revised edition Addison-Wesley Publishing Co. (1968)
34. Gouëzel, S.: Central limit theorem and stable laws for intermittent maps. *Probab. Theory Related Fields* **128** (2004), no. 1, 82–122
35. Gouëzel, S.: Sharp polynomial estimates for the decay of correlations. *Israel J. Math.* **139** (2004), 29–65
36. Gouëzel, S.: Berry-Esseen theorem and local limit theorem for non uniformly expanding maps. *Ann. Inst. H. Poincaré Probab. Statist.* **41** (2005), no. 6, 997–1024
37. Gouëzel S.: Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probab.* **38** (2010), no. 4, 16391671
38. Hall, P., Heyde, C. C.: *Martingale limit theory and its application*. Probability and Mathematical Statistics. Academic Press (1980)
39. *Handbook of dynamical systems*. Vol. 1A. Edited by B. Hasselblatt and A. Katok. North-Holland, Amsterdam, 2002
40. *Handbook of dynamical systems*. Vol. 1B. Edited by B. Hasselblatt and A. Katok. Elsevier B. V., Amsterdam, 2006
41. Hennion, H., Hervé, L.: *Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness*. Lecture Notes in Mathematics **1766**. Springer-Verlag, Berlin, 2001
42. *The theory of chaotic attractors*. Dedicated to James A. Yorke in commemoration of his 60th birthday. Edited by Brian R. Hunt, Judy A. Kennedy, Tien-Yien Li and Helena E. Nusse. Springer-Verlag, New York, 2004
43. Kachurovskii, A. G.: Rates of convergence in ergodic theorems. *Russian Math. Surveys* **51** (1996), no. 4, 653–703
44. Katok, A., Hasselblatt, B.: *Introduction to the modern theory of dynamical systems*. Encyclopedia of Mathematics and its Applications **54**. Cambridge University Press, Cambridge, 1995
45. Kifer, Y.: Nonconventional limit theorems. *Probab. Theory Related Fields* **148** (2010), no. 1-2, 71–106
46. Krengel, U.: *Ergodic theorems*. De Gruyter Studies in Mathematics **6**. Walter de Gruyter, Berlin, 1985
47. Lacey, M., Philipp, W.: A note on the almost sure central limit theorem. *Statist. Probab. Lett.* **9** (1990), no. 3, 201–205
48. Lalley, S., Nobel, A. B.: Denoising deterministic time series. *Dyn. Partial Differ. Equ.* **3** (2006), no. 4, 259279
49. Ledoux, M.: *The concentration of measure phenomenon*. Mathematical Surveys and Monographs **89**. American Mathematical Society, Providence, RI, 2001
50. Melbourne, I.: Large and moderate deviations for slowly mixing dynamical systems. *Proc. Amer. Math. Soc.* **137** (2009), no. 5, 1735–1741
51. Melbourne, I., Nicol, M.: Large deviations for nonuniformly hyperbolic systems. *Trans. Amer. Math. Soc.* **360** (2008), no. 12, 6661–6676
52. Melbourne, I., Nicol, M.: A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Ann. Probab.* **37** (2009), no. 2, 478505
53. Philipp, W., Stout, W.: Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* **2** (1975), issue 2, no. 161

- 54. Rey-Bellet, L., Young, L.-S.: Large deviations in non-uniformly hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems* **28** (2008), no. 2, 587–612
- 55. Sarig, O.: Subexponential decay of correlations. (English summary) *Invent. Math.* **150** (2002), no. 3, 629–653
- 56. Young, L.-S.: Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)* **147** (1998), no. 3, 585–650
- 57. Young, L.-S.: Recurrence times and rates of mixing. *Israel J. Math.* **110** (1999), 153–188
- 58. Young, L.-S.: What are SRB measures, and which dynamical systems have them? Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. *J. Statist. Phys.* **108** (2002), no. 5–6, 733–754
- 59. Young, L.-S.: Decay of correlations for certain quadratic maps. *Comm. Math. Phys.* **146** (1992), no. 1, 123–138